ARCHIVES OF ACOUSTICS
32, 3, 603–616 (2007)

DETECTION AND RECOGNITION OF NATURAL SOUNDS

Kim ABOUCHACRA[(1)], Tomasz ŁĘTOWSKI, Jessica GOTHIE

U.S. Army Research Laboratory
Human Research and Engineering Directorate
Aberdeen Proving Ground, MD 21005-5425
e-mail: trl2@arl.army.mil

(*received May 22, 2007; accepted June 18, 2007*)

Human-machine interfaces and environment simulators increasingly depend on audio interfaces. Acoustic signals are used to provide warnings, feedback, information about the state of a system, and to enhance the immersive character of virtual reality environments. In order to decrease the mental workload of the listener, increase the speed of interaction, and minimize the chances for operational error, the audio signals (auditory icons) should have a natural character and clearly differ in their spatial, spectral, and temporal characteristics. Therefore, the design and selection of audio signals for specific applications should be based on the detectability and recognizability of the signals in the intended environments and on the meaningful connotations of the individual sounds. The present study was conducted to assess the detection and recognition thresholds of 30 pre-selected sounds and to determine the specific acoustic properties that make complex natural sounds effective auditory icons. The results of the study revealed a strong dependence of both types of threshold on the type of sound and a relative independence of both thresholds. The sound level difference between the detection and recognition thresholds varied from 1 to 13 dB and should be considered as an important criterion in auditory icon selection.

Keywords: natural sound, detection, recognition.

1. Introduction

Natural sounds produced by various noisemakers, such as bells and rattles, have been the traditional signals used in hearing tests for young children. Their main disadvantage is that they are not frequency specific and thus have a limited diagnostic value. In the late 1960s, RAKOWSKI and ŁĘTOWSKI [1] developed the concept of tonal audiometry based on octave-filtered sound effects and recorded the first such test for the Institute for the Mother and Child in Warsaw. The authors took advantage of the formant

[(1)] Presently with the Department of Otolaryngology – Head and Neck Surgery, American University of Beirut, Phase I – 6-th Floor, Beirut 1107 2020; e-mail: ks05@aub.edu.lb.

structure of many natural sounds and used octave-band filtered sounds as the frequency-specific simplified versions of natural sounds for hearing test purposes. The test had two versions, one for children 0.5–3 years old and one for children 3–6 years old, and was based on 24 natural sounds that were octave-filtered and centered at octave frequencies extending from 125 Hz to 16,000 Hz. This early work was later expanded on in two related research areas: timbre solfege training and filtered sound-effect hearing tests.

Timbre solfege training is based on using the pitch scale as a tool to teach listeners to hear and differentiate subtle changes in sound timbre. The initial part of the training utilizes narrow-band filtered noises and sound effects but then progresses into more and more subtle changes in the spectral content of a sound. The original training program was developed by RAKOWSKI, SZLIFIRSKI, and ŁĘTOWSKI [2–4] and has been taught at the Fryderyk Chopin Academy of Music in Warsaw since the early 1970s. Versions of this program have been implemented at various times at diverse institutions, such as McGill University, the University of Surrey, General Motors Technical Education Center, and the U.S. Army Research Laboratory.

The initial use of filtered sound-effects for hearing testing by ŁĘTOWSKI and RAKOWSKI [1] resulted in several research studies at the Pennsylvania State University, Montclair State University, and the U.S. Army Research Laboratory, and in a CD version of a new test available from the U.S. Army Research Laboratory [5, 6]. The CD version of the filtered sound-effects test is based on a study conducted by MYERS, ŁĘTOWSKI, ABOUCHACRA, KALB, and HAAS [5] in which a corpus of 25 octave-filtered sound effects was evaluated to determine the most robust sound-database for use in tonal sound effect detection and recognition tasks.

In the process of selecting sounds for the MYERS' *et al.* [5] study, a large number of natural sounds was evaluated, resulting in a starting set of 33 sounds. Further evaluation, though, revealed three of the sounds as not being sufficiently recognizable across a variety of populations, and so they were removed. Later another five less robust sounds were dropped to make the final study manageable. However, before the selected 25 sounds were octave-filtered and initially evaluated, all 30 sounds were used in a preliminary study to determine thresholds of detection and recognition for unfiltered sounds and to evaluate the testing procedure to be used in the main study. The data were collected, used in determining the proper filtering strategies for the final sounds, and left in an archive as an internal reference for the main study.

Over the last 10–15 years there has been a growing interest in studying natural sounds for various medical, communication, and military applications. It has been recognized that the human auditory system may be the best sensory system for performing detection and recognition tasks in uncertain complex environments [7]. Therefore, the auditory modality is becoming the primary modality for providing warning and tactical signals in such environments. Examples of such environments include aircraft cockpits, hospital operation rooms, ship captain's bridges, and nuclear power plant control centers. All of these environments require the control and monitoring of numerous subsystems working in tandem for the safe and proper operation of the whole system.

Several studies have demonstrated that natural sounds are preferable to both speech and synthetic signals as warning and human-machine communication signals. The more natural, easy to hear, and different from one another these signals are, the faster and more reliable is the response of human operators [8, 9]. There is also a growing literature on the selection of natural sounds based on their connotation [10–12] and on applying signal detection theory to signal level selection [13, 14]. In addition, the growing interest in the selection and controlled reproduction of natural sounds is a result of the increasing role of auditory displays in virtual environments. Nonetheless, despite the broad military, medical, and industrial applications of natural sounds, there is a scarcity of data regarding the actual detection and recognition thresholds of natural sounds in various environments. There have been some studies reported at conferences or completed as parts of graduate work that have referred to such data, but it is very hard to find specific data in the open literature [15–19]. Therefore, in order to partially fill this gap, the authors decided to publish their data on the detection and recognition of natural sounds in quiet and noisy environments; data which were collected more than 10 years ago. The goal of the original study was to identify sounds that were easy to recognize as soon as they are detected for the purposes of audiometric testing and as warning and communication signals. The same applications are considered in the present paper.

2. Methodology

2.1. Sounds

A group of 30 natural sounds was used as the target sounds. The sounds were chosen on the basis of their common occurrence and clear association with a sound source. The selected sounds are listed in Table 1 and their spectra are shown in Fig. 1. The sounds were 2.0 to 4.0 seconds in duration and the duration of a sound was determined by the natural time period needed for sound identification. For example, the sound of a dog barking consisted of two short barks and one longer one, together making a complete event that was identifiable as a dog bark. Such complete groups of sounds will be referred to as *environmental events*. The target sounds used in this study either had the form of one long sound (sounds 1, 4, 7–8, 15, 17–20, 22, 24, 27, and 30) or consisted of two (sounds 9–10, 16, and 23), three (sounds 2–3, 5–6, 11–14, and 28), or four (sounds 21 and 29) separate sounds forming an environmental event. Obviously, within each of these categories the sounds differed in their internal structure (internal modulation) and some of the multi-sound events were actually groups of several finer events (e.g., clock chime, cricket, cuckoo). The background noise used in the study was a 20-voice multitalker noise (MTN) [20] presented at 60 dB A. This noise has a relatively flat spectrum from 200 to 1000 Hz (with a slight peak around 600 Hz) and decreases thereafter with a 6 dB/oct slope.

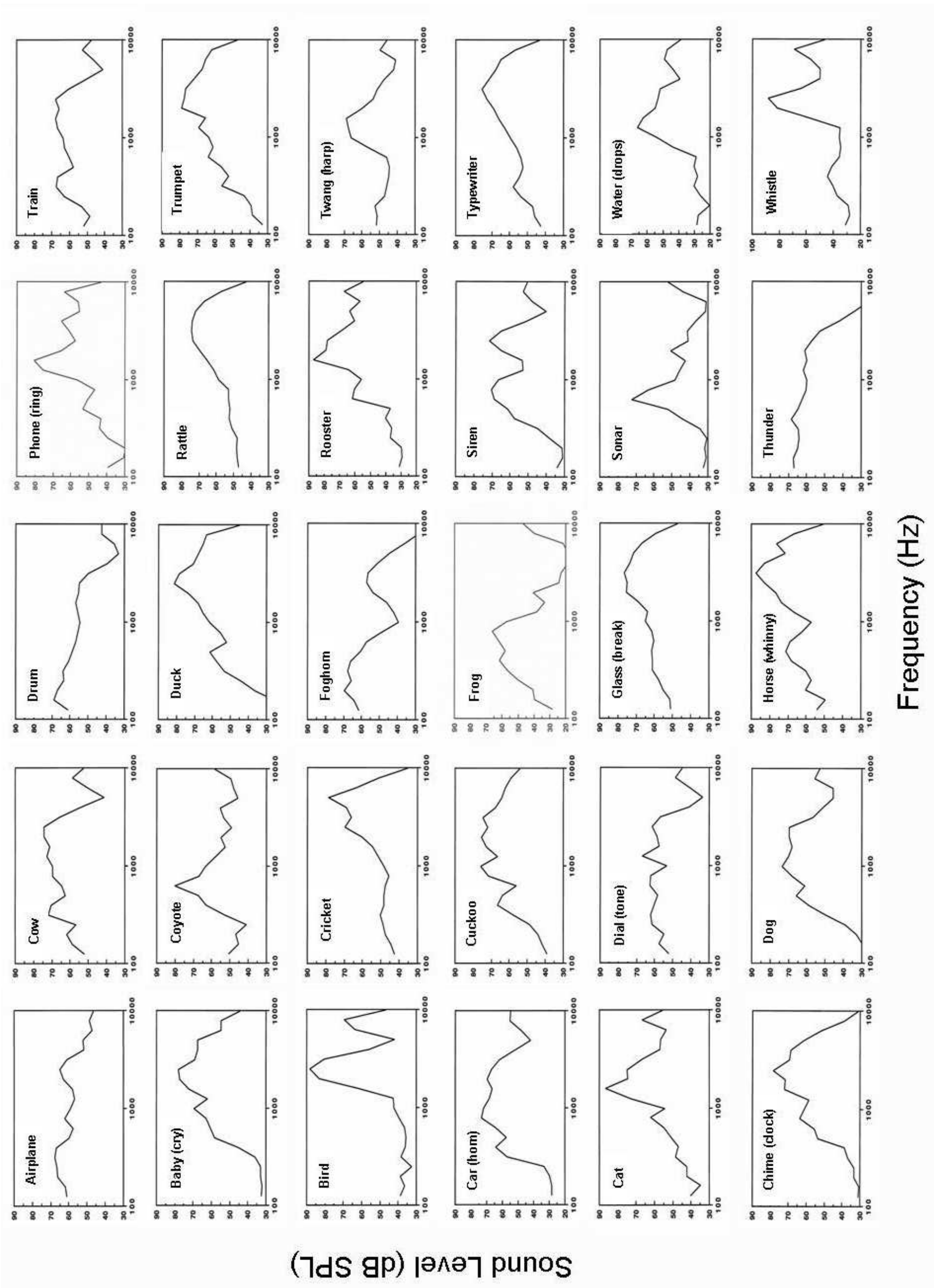

Fig. 1. Spectral characteristics of the 30 sounds used in the study.

Table 1. Target sounds used in the study.

1	Airplane	11	Dial (tone)	21	Rooster
2	Baby (cry)	12	Dog	22	Siren
3	Bird	13	Drum	23	Sonar
4	Car (horn)	14	Duck	24	Thunder
5	Cat	15	Foghorn	25	Train
6	Chime (clock)	16	Frog	26	Trumpet
7	Cow	17	Glass (break)	27	Twang (harp)
8	Coyote (howl)	18	Horse (whinny)	28	Typewriter
9	Cricket	19	Phone (ring)	29	Water (drops)
10	Cuckoo (clock)	20	Rattle	30	Whistle

2.2. Listeners

A group of 20 listeners, 18 to 43 years of age (15 females and 5 males; mean age 25.3 years, SD of 7.5 years) participated in the study. All listeners had left and right ear hearing thresholds of 20 dB HL or better, measured at octave frequencies in the 250 to 4000 Hz audiometric range. The listeners were the same subjects who participated in the companion study by MYERS *et al*. [5]. Their average hearing thresholds in dB SPL measured at octave frequencies from 250 to 8000 Hz are listed in Table 2.

Table 2. Average pure tone hearing thresholds (M) in dB SPL and their standard deviations (SD) in dB of the group of 20 listeners participating in the study.

Frequency (kHz)	0.25	0.50	1.00	2.00	4.00	8.00
M (dB SPL)	18.8	12.6	8.9	12.9	10.8	17.2
SD (dB)	7.2	5.8	5.3	5.8	6.7	8.6

2.3. Instrumentation

All testing was conducted in an audiometric booth with ambient noise levels suitable for sound-field testing [21]. The sounds were presented monaurally using an ER-1 insert earphone equipped with a TIP-50 eartip. The better ear was selected for signal delivery and the contralateral ear was occluded with an EAR™ foam earplug. All target sounds were digitized and played from a PC computer using a Tucker-Davis Technologies (TDT) System II signal control system and proprietary software. The multitalker noise was played from a Nakamishi MR-1 cassette player and mixed together with target sounds for earphone presentation. The noise level was 60 dB A, measured in an Occluded Ear Simulator (Zwislocki Coupler) [22]. All the instrumentation and calibration procedures were identical to those used by MYERS *et al.* [5].

2.4. Procedure

Prior to data collection all the sounds were presented to the listeners at a comfortable listening level of 40 dB HL and identified by their names. After a short break, the same sounds were presented in a different order and the listeners' task was to identify the sounds. Eighteen listeners correctly identified all 30 natural sound upon one presentation. The remaining two listeners correctly identified 28 of the sounds after one presentation but required a second presentation to correctly identify the two other sounds.

During the study all listeners completed two listening tasks: a detection task and a recognition task. Both tasks were completed in two listening conditions: in quiet (Q) and in noise (MTN).

Detection Task: Detection thresholds were measured using the "best PEST" adaptive variant of the Yes-No procedure [23, 24]. The procedure was set to estimate the 50% detection point on the psychometric curve. All 30 sounds were tested in random order and a random selection of three of the sounds was retested for each listener to assess repeatability of the data.

Recognition Task: Recognition thresholds were measured using an ascending variant of the method of limits. At the beginning of the procedure the sound level of each target signal was set at 6 dB level below the listener's detection threshold for the signal, and the level was gradually increased in 2 dB steps. At each signal presentation level the listener's task was to identify the presented sound by selecting one of the 30 available responses on a computer screen. The procedure continued until the listener had correctly identified the sound at three consecutive presentation levels. The first of these three levels was deemed the recognition threshold for the given sound.

The detection task was always presented before the recognition task for the same condition (Q or MTN) and both these tasks were performed in direct succession. The order of the quiet and noise conditions was counterbalanced. The order of the sound effects was randomized for each task, condition, and listener. Further details regarding the testing procedure can be found in MYERS *et al.* [5].

3. Results and discussion

3.1. Detection thresholds

Average detection thresholds and their respective standard deviations for each of the 30 environmental sounds used in this study are listed in Table 3. The detection thresholds for individual sounds varied from 11.2 (sonar) to 27.8 (airplane) dB SPL in quiet and from 34.2 (phone) to 48.9 (rattle) dB SPL in 60 dB A multi-talker noise.

Table 3. Detection and recognition thresholds for 30 environmental sounds presented in quiet (Q) and in the background of a 60 dB A multi-talker noise (MTN). Mean – arithmetic mean in dB SPL; SD – standard deviation in dB.

Sound number	Target sound	Detection threshold				Recognition threshold			
		Quiet (Q)		Noise (MTN)		Quiet (Q)		Noise (MTN)	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD
1	Airplane	27.8	5.0	46.2	8.8	36.4	7.3	52.9	7.2
2	Baby (cry)	16.6	5.0	41.6	5.4	21.9	9.7	48.9	6.8
3	Bird	16.3	8.1	45.8	8.6	20.1	10.1	49.2	6.0
4	Car (horn)	17.2	5.9	34.6	5.7	20.5	9.8	39.8	8.0
5	Cat	16.0	6.2	37.8	8.6	19.9	9.6	45.6	7.8
6	Clock (chime)	22.0	7.1	43.4	12.0	26.9	7.9	52.2	10.5
7	Cow	22.7	8.1	40.5	10.5	33.9	8.3	50.2	10.3
8	Coyote (howl)	16.4	6.7	42.1	6.0	21.6	9.4	42.2	5.6
9	Cricket	21.4	6.9	43.5	5.9	24.5	7.3	47.4	8.7
10	Cuckoo	19.0	6.3	43.3	6.3	27.3	7.8	48.8	7.8
11	Dial (tone)	14.8	7.4	47.4	5.7	24.6	10.9	48.8	10.3
12	Dog	17.3	7.5	40.5	9.9	24.3	11.2	50.6	7.8
13	Drum	21.0	5.4	37.7	5.3	35.7	7.1	46.2	6.2
14	Duck	23.7	8.6	43.1	6.3	29.2	10.7	51.5	7.7
15	Foghorn	15.1	8.1	43.3	5.0	29.9	10.3	52.6	7.1
16	Frog	15.4	4.8	40.1	6.1	28.0	7.2	51.6	7.0
17	Glass	21.1	8.8	38.8	6.7	30.2	7.2	48.3	6.3
18	Horse (whine)	13.0	5.9	41.9	5.1	16.0	8.9	47.6	9.8
19	Phone (ring)	12.0	6.9	34.2	7.6	16.7	9.6	42.0	8.7
20	Rattle	21.9	6.3	48.9	13.6	29.5	12.4	56.5	12.8
21	Rooster	17.7	7.0	39.2	7.8	23.8	8.4	45.9	7.5
22	Siren	18.3	10.7	38.0	7.4	23.5	9.9	47.7	9.8
23	Sonar	11.2	13.4	45.0	8.8	16.6	12.4	50.3	8.6
24	Thunder	25.2	9.8	47.0	4.9	35.3	11.1	57.0	7.9
25	Train	22.5	7.4	47.4	6.8	31.0	11.2	49.8	7.8
26	Trumpet	19.5	5.6	44.7	8.3	25.4	9.0	49.7	4.9
27	Twang (harp)	23.2	5.8	44.1	10.9	27.8	6.8	46.5	10.5
28	Typewriter	23.7	9.0	46.0	9.0	29.3	10.8	46.9	8.8
29	Water (drops)	26.2	5.8	41.5	4.9	37.8	8.8	50.5	7.3
30	Whistle	16.8	5.7	38.6	9.1	19.2	6.4	45.4	8.4

The latter range is equivalent to a -25.8 to -11.1 dB range of signal-to-noise ratios (SNR) determined as the ratio of the signal level in dB SPL to the noise level in dB A. Mean detection thresholds calculated across the sounds were 19.5 and 42.2 dB SPL for quiet and noise conditions, respectively. The latter value corresponds to a -17.8 dB SNR (for the 60 dB A MTN).

The range of the detection thresholds reported in Table 3 spans 16.6 dB for quiet and 14.7 dB for noise condition. The only other study where similar data were presented was the study by WATSON, KIDD, and GYGI [17] quoted by GYGI [19] in which he reported a slightly narrower range of 12 dB, extending from -20 to -8 SNR for the detection of natural sounds in white noise. However, he also reported that except for two outlying sounds, the detection thresholds for the sounds ($n = 25$) used in the study were all within 7 dB of each other. Likewise, in the present study, the detection thresholds for all but three of the sounds fall within an 11 dB range for both the quiet and noise conditions. These data seem to imply that that the range of variability in detecting most sounds in a given listening condition is about 10 dB except for very low frequency sounds that require higher thresholds. Most of the data for filtered sound effects reported by MYERS *et al.* [5] also supports this observation.

Standard deviation of the mean detection thresholds ranged from 4.8 to 13.4 dB in quiet and from 4.9 to 13.6 dB in noise. Overall, these values are relatively similar to the standard deviations reported by MYERS *et al.* [5], although there are some larger individual differences for several of the filtered sound effects and natural sounds.

The easiest sounds to detect were in general characterized by one or two strong formants in the middle frequency range (e.g., sounds of bird, cat, coyote, frog, horse, phone, sonar, whistle). Two other cues facilitating sound detectability were an abrupt beginning and/or end to the sound. The relatively easy detection of some high-frequency sounds in noise was most likely due to the relatively low level of the masking noise (60 dB A) and its lack of high-frequency energy. The sounds that were the most difficult to detect in both quiet and noise were sounds with a dominant broad concentration of energy in the low frequency range (e.g., airplane, thunder, train). Several high frequency sounds were relatively easy to detect in quiet but much harder to detect in noise (e.g., bird, rattle).

It must also be stated that the detection thresholds for natural sounds reported in this study and also in those by GYGI *et al.*, [17, 19] are much lower than the detection thresholds for natural sounds reported elsewhere in the sparse literature. For example, OLLERHEAD [25] reported that the helicopter blade slap can be detected at a -5 dB SNR with respect to the surrounding noise level and DOLL and HANNA [26] reported -3 dB detection threshold for a simulated sonar signal presented in white noise. Reasons for these differences are unclear.

3.2. Recognition thresholds

Average recognition thresholds and their standard deviations are listed together with the detection thresholds in Table 3. Individual recognition thresholds vary from 16.0

(horse whinny) to 37.8 (water drops) dB SPL and from 39.8 (car horn) to 57.0 (thunder) dB SPL in quiet and noise, respectively. Mean recognition thresholds calculated across all the sounds were 25.9 for quiet and 48.8 dB SPL for noise. The latter value corresponds to a -11.2 dB SNR (for the 60 dB A MTN). The respective standard deviations for sound recognition in quiet and in noise varied from 6.4 to 12.4 dB and from 5.6 to 12.8 dB. These values represent similar ranges of standard deviations to those reported by MYERS *et al.* [5].

One reference point that can be used as a comparison to the recognition thresholds for natural sounds is the speech reception threshold. It is generally assumed that speech is understood at about 10 dB SPL (digits) to 20 dB SPL (spondees) in quiet and at -4 to -12 dB SNR in noise [27–32]. MILLER *et al.* [33] and O'NEIL [34] also reported that the recognition threshold for isolated words is about 6 to 7 dB higher than for sentences.

The overall variability of the recognition threshold measured in this study under quiet and noise conditions was equal to 21.8 and 17.2 dB, respectively. These values indicate the expected variability range for natural sound recognition in general to be about 20 dB. Verbal reports by the listeners imply that the overall temporal envelope and the dominant frequency (pitch) were important features that they relied on for sound detection. Then, as the sound level approached the recognition level, the listeners relied more and more on sound character (spectrum) and internal modulation.

The SNRs needed for sound recognition in noise varied from -20.2 (carhorn) to -3.0 (thunder) dB and the low frequency sounds dominated the group of sounds that were the hardest to recognize. The reported values represent a similar range to the range of recognition thresholds for natural sounds in white noise reported by GYGI [17, 19], although the range determined in his study was shifted downwards and extended from -30 to – 8 dB.

3.3. Recognition-detection gap

The recognition thresholds of the individual sounds exceeded detection thresholds by 1.4 to 12.6 dB, and in general, these differences had a similar range in quiet and in noise. The sound level differences between the recognition and detection thresholds for the individual sounds in both quiet and noise are shown in Fig. 2. Comparisons between recognition-detection gaps in quiet and noise for most of the sounds ($n = 26$) were within -2.8 to $+4.5$ dB of each other, with most of the sounds showing a slightly larger gap in noise. Four sounds that had a very small recognition-detection gap in noise produced slightly larger quiet vs. noise gap differences exceeding those listed above. These sounds had overall spectral (e.g., dial, train) and temporal (e.g., train, typewriter) properties that quite closely resembled those of the multi-talker noise. Some other sounds that show small recognition-detection gap in noise (e.g., coyote) had the main peak of their energy coinciding with the frequency of the energy peak of the multi-talker noise spectrum. However, both larger and smaller recognition-detection gaps in noise can result from sound-noise interactions. In sum, the obtained results support poor overall correlation between the detection and recognition thresholds observed in MYERS *et al.* [5] and in other studies [17, 19].

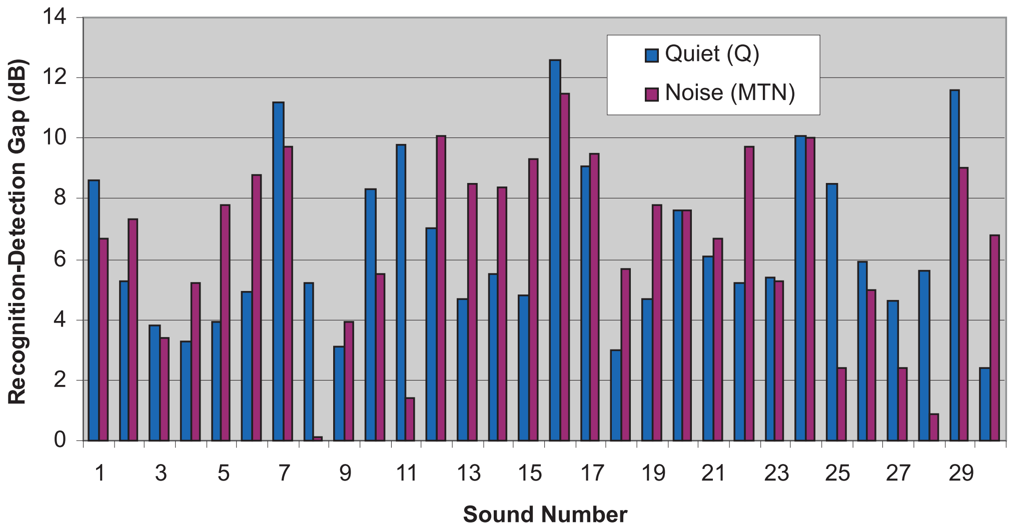

Fig. 2. Recognition-detection gaps in quiet (Q) and in multi-talker noise (MTN). Sound numbers correspond to the numbers used in Table 1.

The differences between detection and recognition thresholds obtained in this study were twice as small as those reported by MYERS *et al.* [5] for filtered sound effects. This should be expected given the greater number of potential cues that unfiltered sounds provide to the listener, which then facilitate the task of sound recognition.

3.4. Data repeatability

The retest trials presented at the end of each test block were used to spot check the repeatability of the listeners' responses. A random selection of three sounds was presented in each of the basic test conditions resulting in 60 retest trials (20 listeners $\times 3$ sounds) under each listening condition (quiet and noise). All retest data for the detection threshold in quiet were within -9 to $+8$ dB of the test data. In addition, most of the retested thresholds ($n = 39$) were ± 4 dB with respect to the original thresholds. The retest data for the detection threshold in noise were within -7 to $+9$ dB of the test data. Similarly to detection in quiet, more than 50% of the retest responses ($n = 32$) were ± 4 dB relative to the test data. Therefore, in both the above cases, the retest data demonstrate relatively good repeatability of the obtained thresholds. This observation supports the notion that the variability observed in the reported data is mainly due to the differences between the sounds and the listeners and to a lesser degree due to listener inconsistency.

3.5. Common errors

An analysis of the errors made by listeners in the sound recognition task revealed that the listeners tended to confuse sounds having similar acoustic properties. Sounds that were confused in both quiet and noise typically had very similar spectra (e.g. baby-cat-rooster, drum-thunder-airplane-foghorn, glass-horse, rattle-typewriter) or temporal patterns (e.g. cuckoo-rooster, water-train, drum-twangs). The frequency of these errors was similar in quiet and in noise and there were no systematic differences among the listeners.

However, the listeners did demonstrate systematic preferences in selecting some sounds more often than others under conditions of uncertainty. The most frequently "guessed" sounds were airplane, cricket, typewriter, and bird, whereas "guesses" that were relatively seldom used were frog, trumpet, dog, and cow. An analysis of the naming errors made by the listeners seems to indicate that several listeners established internal preferences (standards) for low-frequency sounds (e.g., airplane), high-frequency sounds (e.g., cricket), slowly-changing sounds (e.g., typewriter), and fast-changing sounds (e.g., bird) and used these standards in case of uncertainty. Such systematic preferences also imply that high-low and slow-fast dichotomies are the main criteria in sound recognition.

4. Conclusions

Despite the availability of extensive literature on signal detection there is a marked scarcity of data on the detection of natural sounds. Almost all of the studies published in open literature are limited to pure tones, narrow bands of noise, and speech signals. The main reason for this situation is the huge variety of natural sounds and acoustic environments, which does not lend itself to the development of comprehensive psychoacoustic measures. Yet, there is a need for a psychoacoustic database that could be used as a guide in selecting natural sounds for specific applications.

The detection thresholds obtained in this study indicate that most natural sounds are detectable at 10 to 20 dB SPL in quiet environments. When the same sounds are presented in a 60 dB A multi-talker noise, they are usually detectable at a -15 dB SNR, although some sounds can be detected at levels as low as -20 dB, while others are not detectable until they reach about a -10 dB SNR.

Reported data show that the recognition thresholds vary among sounds more than their detection thresholds. Sound recognition may require a signal that is more than 10 dB higher than its detection threshold (e.g., rattle, thunder), although some sounds are recognizable almost as soon as they are detected (e.g., cricket, horse). In reported study, most of the observed recognition-detection gaps were in the 5 to 10 dB range. For many sounds these gaps were relatively similar in quiet and in noise. However, the thresholds obtained in noise were undoubtedly affected by the noise selection and may be different in other types of noise. It must be also stressed that specific sound-noise interactions can either decrease or increase the size of the recognition-detection gap observed in quiet. Therefore the observed variability of the recognition-detection gap data can be considered as experimental evidence for the absence of a simple relation between the detection and recognition processes for natural sounds. In other words, presented data do not provide support for the generalization of the detection-recognition theorem [29, 30] on the whole domain of natural sounds.

Some specific sound characteristics that were reported by the listeners as helping them in sound detection included abrupt onset and/or offset of sound, and strong formants in the middle frequency range. Sound recognition was later aided by specific temporal modulation within the sound (either in amplitude or in the frequency domain) and specific timbral (spectral) properties of the sound.

A comparison of threshold differences between recognition and detection of natural sounds indicate that sounds dominated by either low-frequency or high-frequency energy were harder to recognize than those dominated by middle-frequency energy and usually resulted in large recognition-detection gaps. Therefore, such sounds are not good candidates for warning and tactical signals in multi-signal environments. Conversely, it seems reasonable to postulate that the sounds having small and similar in both quiet and noise recognition-detection gaps are good candidates for robust auditory icons and informational signals that can be effectively used in a specific family of environments.

In closing, it should be stressed that presented data are in basic agreement with the data published by Myers *et al.* [5] for a filtered subset of the same sounds. This agreement confirms that the filtered sounds used in the Myers *et al.* study preserved the basic character of the unfiltered sounds and can be treated as faithful frequency-specific icons representing the original natural sounds.

References

[1] Łętowski T., Rakowski A., *Environmental audiometry: Filtered sound effects for testing hearing in children* [in Polish], Audiometric test developed for the Institute for the Mother and Child, AMiFC, Warszawa 1970.

[2] Rakowski A., Łętowski T., Szlifirski K., Okoń B., *Methods for developing sensitivity to timbre in music students* [in Polish], [in:] Studies in Sound Transmission Theory, Gryzik A., Urbański J., Wołk–Łaniewski N. [Eds.], Wydawnictwa RiTV, pp. 183–194. Warszawa 1982.

[3] Łętowski T., *Development of technical listening skills: Timbre Solfeggio*, Journal of the Audio Engineering Society, **33**, 240–244 (1985).

[4] Miśkiewicz A., *Timbre Solfege: A course in technical listening for sound engineers*, Journal of the Audio Engineering Society, **40**, 621–625 (1992).

[5] Myers L., Łętowski T., Abouchacra K., Kalb J., Haas E., *Detection and recognition of octave-band sound effects*, Journal of the American Academy of Audiology, **7**, 346–357 (1996).

[6] Abouchacra K., Łętowski T., *Comparison of air-conduction and bone-conduction hearing thresholds for pure tones and octave-band filtered sound effects*, Journal of the American Academy of Audiology, **10**, 422–428 (1999).

[7] KANTOWITZ BARRY H., SORKIN RD, *Human factors: Understanding people-system relationships*, John Wiley & Sons, New York 1983.

[8] BALLAS J., *Acoustic and perceptual-cognitive factors in the identification of 41 environmental sounds*, Office of Naval Research, Technical Report, ONR-89-2, 1–84 (1989).

[9] BALLAS J. A., *Common factors in the identification of an assortment of brief everyday sounds*, Journal of Experimental Psychology: Human Perception & Performance, **19**, 250–267 (1993).

[10] GAWER W., *Auditory icons: Using sound in computer interfaces*, Human-Computer Interactions, **2**, 167–177 (1986).

[11] GYGI B., KIDD G. R., WATSON C. S., *Spectral-temporal factors in the identification of environmental sounds*, Journal of the Acoustical Society of America, **15**, 1252–1265 (2004).

[12] SHAFIRO V., GYGI B., *How to select stimuli for environmental sound research and where to find them?*, Behavior Research Methods, Instruments & Computers, **36**, 590–598 (2004).

[13] FIDELL S., PEARSONS K., BENNETT R., *Prediction of aural detectability of noise signals*, Human Factors, **16**, 373–383 (1974).

[14] RAPOZA A., FLEMING G., *Determination of a sound level for railroad horn regulatory compliance*, U.S. Department of Transportation, Federal Railroad Administration, Technical Report DTS-34-RR397-LR1, 2002.

[15] ABOUCHACRA K., MUNDIS P., HAAS E., ŁĘTOWSKI T., MYERS L., *Detection and recognition of familiar sounds*, Journal of the Acoustical Society of America, **95**, 2967(A) (1994).

[16] STELLMACK M., LUFTI R., *Detection and recognition of familiar sounds in noise*, Journal of the Acoustical Society of America, **97**, 2967(A) (1995).

[17] WATSON C., KIDD G., GYGI B., *Detection and identification of environmental sounds*, Journal of the Acoustical Society of America, **99**, 1516(A) (1996).

[18] GYGI B, WATSON C. S., KIDD G. R., *Factors in the detection of environmental sounds*, Journal of the Acoustical Society of America, **101**, 3149(A) (1997).

[19] GYGI B., *Factors in the identification of environmental sounds*, Doctoral dissertation completed at Department of Psychology in Indiana University, 2001.

[20] FRANK T., CRAIG C., *Comparison of Auditec and Rintelmann recordings of NU-6 lists*, Journal of Speech and Hearing Disorders, **49**, 267–271 (1984).

[21] American National Standards Institute (ANSI), *Maximum permissible ambient noise levels for audiometric test rooms*, ANSI S3.1-1991, New York 1991.

[22] American National Standards Institute (ANSI), *Occluded ear simulator*, ANSI S3.25-1989 (R2003), New York 1989.

[23] PENTLAND A., *Maximum likelihood estimation: the best PEST*, Perception and Psychophysics, **28**, 377–379 (1980).

[24] LIEBERMAN H., PENTLAND A., *Microcomputer-based estimation of psychophysical thresholds: the best PEST*, Behavioral Research Methods and Instrumentation, **12**, 21–25 (1982).

[25] OLLERHEAD J., *Helicopter aural detectability*, U.S. Army Air Mobility R&D Laboratory, Technical Report 71–33, 1971.

[26] DOLL T., HANNA T., *Enhanced detection with bimodal sonar display*, Human Factors, **31**, 539–550 (1989).

[27] PLOMP R., MIMPEN A., *Speech-reception threshold for sentences as a function of age and noise level*, Journal of the Acoustical Society of America, **66**, 1333–1342 (1979).

[28] PUNCH J., HOWARD M., *Spondee recognition threshold as a function of set size*, Journal of Speech and Hearing Disorders, **50**, 120–125 (1985).

[29] NORDHOEK I., HOUTGAST T., FESTEN J., *Measuring the threshold for speech reception by adaptive variation of the signal bandwidth. I. Normal-hearing listeners*, Journal of the Acoustical Society of America, **105**, 2895–2902 (1999).

[30] NORDHOEK I., HOUTGAST T., FESTEN J., *Measuring the threshold for speech reception by adaptive variation of the signal bandwidth. II. Hearing-impaired listeners*, Journal of the Acoustical Society of America, **107**, 1685–1696 (2000).

[31] VERSFELD N., DAALDER L., FESTEN J., HOUTGAST T., *Method for the selection of sentence materials for different measurement of speech reception threshold*, Journal of the Acoustical Society of America, **107**, 1671–1684 (2000).

[32] RAMKISSOON I., PROCTOR A., LANSING C., BILGER R., *Digit speech recognition thresholds for non-native speakers of English*, American Journal of Audiology, **11**, 1–6 (2002).

[33] MILLER G., HEISEN G., LICHTEN W., *The intelligibility of speech as a function of context of the text material*, Journal of Experimental Psychology, **41**, 329–335 (1951).

[34] O'NEIL J., *Recognition of intelligibility test materials in context and isolation*, Journal of the Speech and Hearing Disorders, **22**, 87–90 (1957).

[35] GREEN D., WEBER D., DUNCAN J., *Detection and recognition of pure tones in noise*, Journal of the Acoustical Society of America, **62**, 948–954 (1977).

[36] WEBER D., *Detection and recognition of auditory patterns*, Perception and Psychophysics, **46**, 1–8 (1989).